Relationship between brain connectivity and human traits

Jiaying Li & Tzu-Chi (Stephanie) Lin

UNC-Chapel Hill

May 8, 2021

Brain Connectome

- The "Connectome" refers to the complex interconnected network of neurons in human brain.
- Brain connectomes use network science to handle the complex methodologies needed to understand the dynamic interactions of different brain regions both functionally and structurally

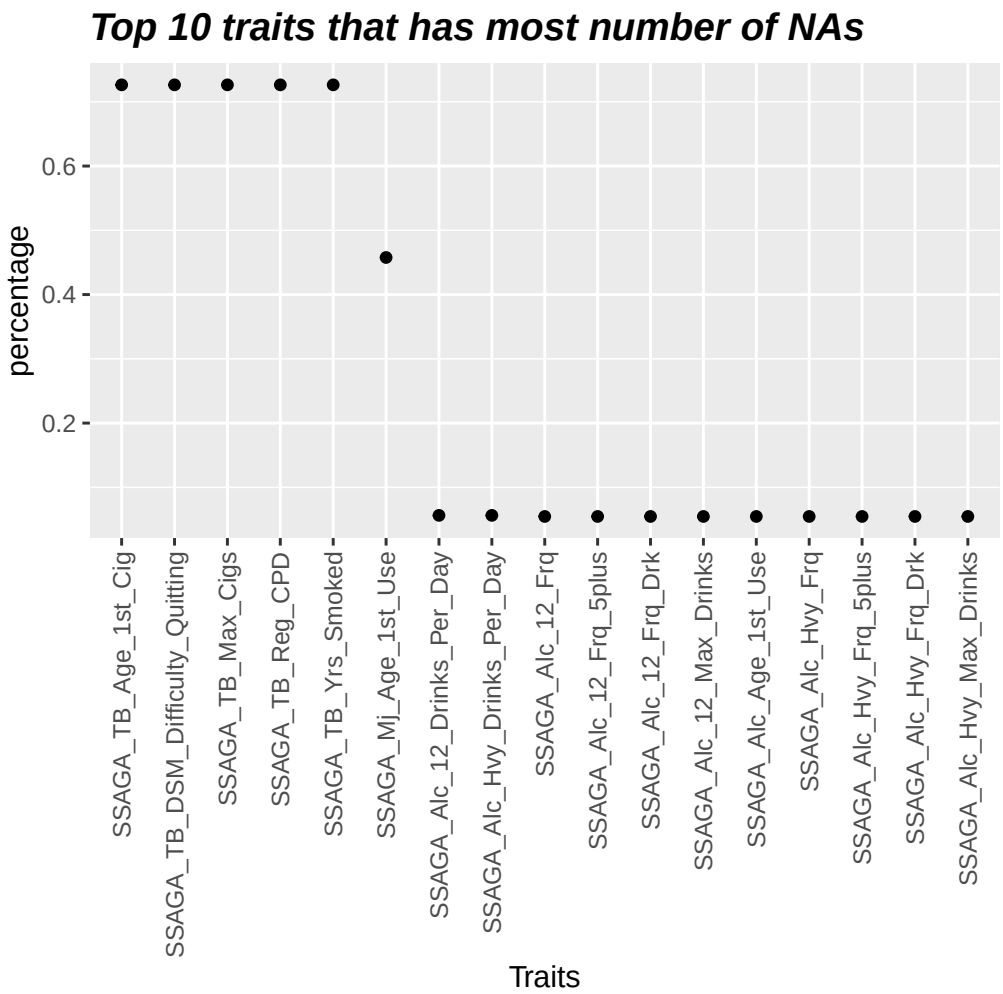
Outline

Outline

Data

- Human Connectome Project:
  - Structural Connectomes (SC)
  - Functional Connectomes (FC)
  - Tensor PC scores for SC
  - Tensor PC scores for FC
- 175 Traits:
  - cognition
  - substance use
  - psychiatric and life function
  - sensory
  - emotion
  - health and family history
  - alertness
  - motor
  - personality

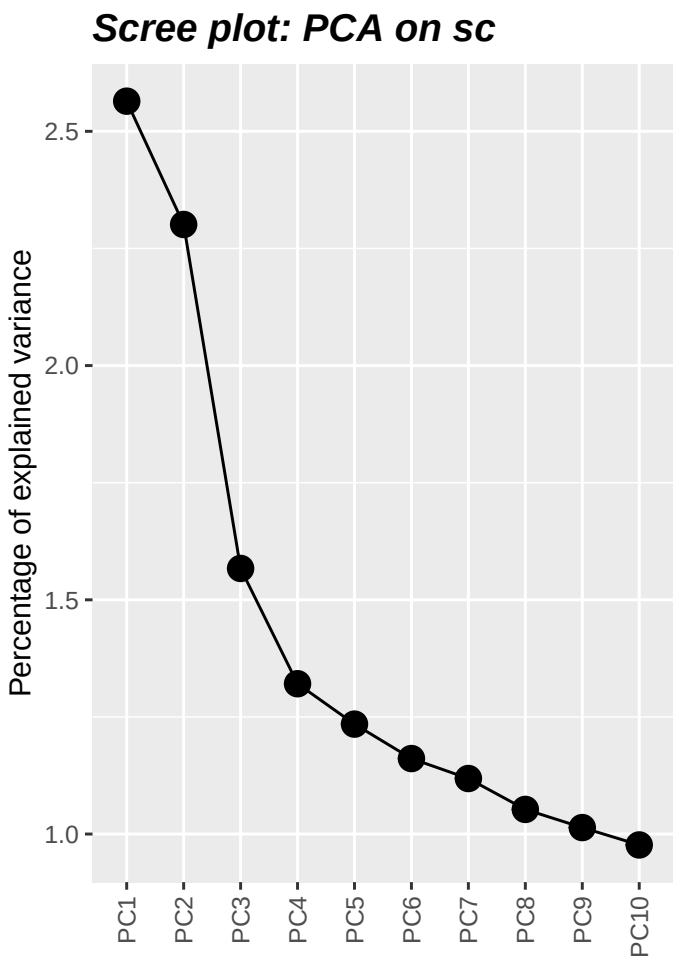
Missing values

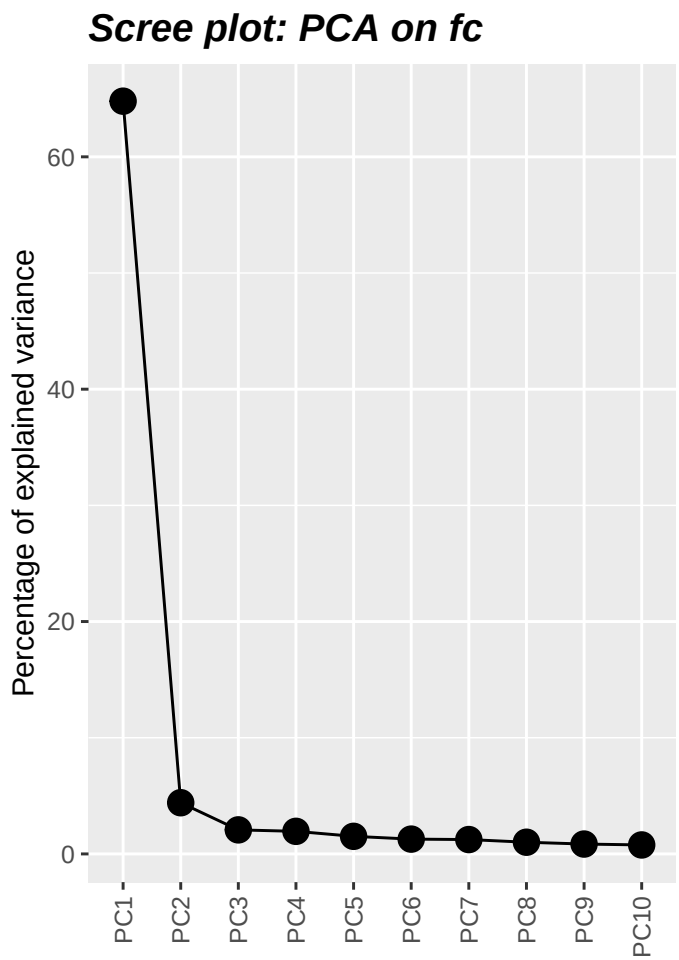


Data management process

- For each subject, vectorize the structural connectomes (SC) and funcational connectomes (FC).
- Remove 76 all zeros columns in vectorized SC.
- Remove 6 substance use traits that contain more than 40% NAs.
- Merge vectorized SC, vectorized FC, TNPCA data, and 169 traits based on subjects.
- Remove 7 subjects without FC information.
- Remove trait that contains only one constant value: "EVA_Num".
- Apply PCA to vectorized SC and FC.

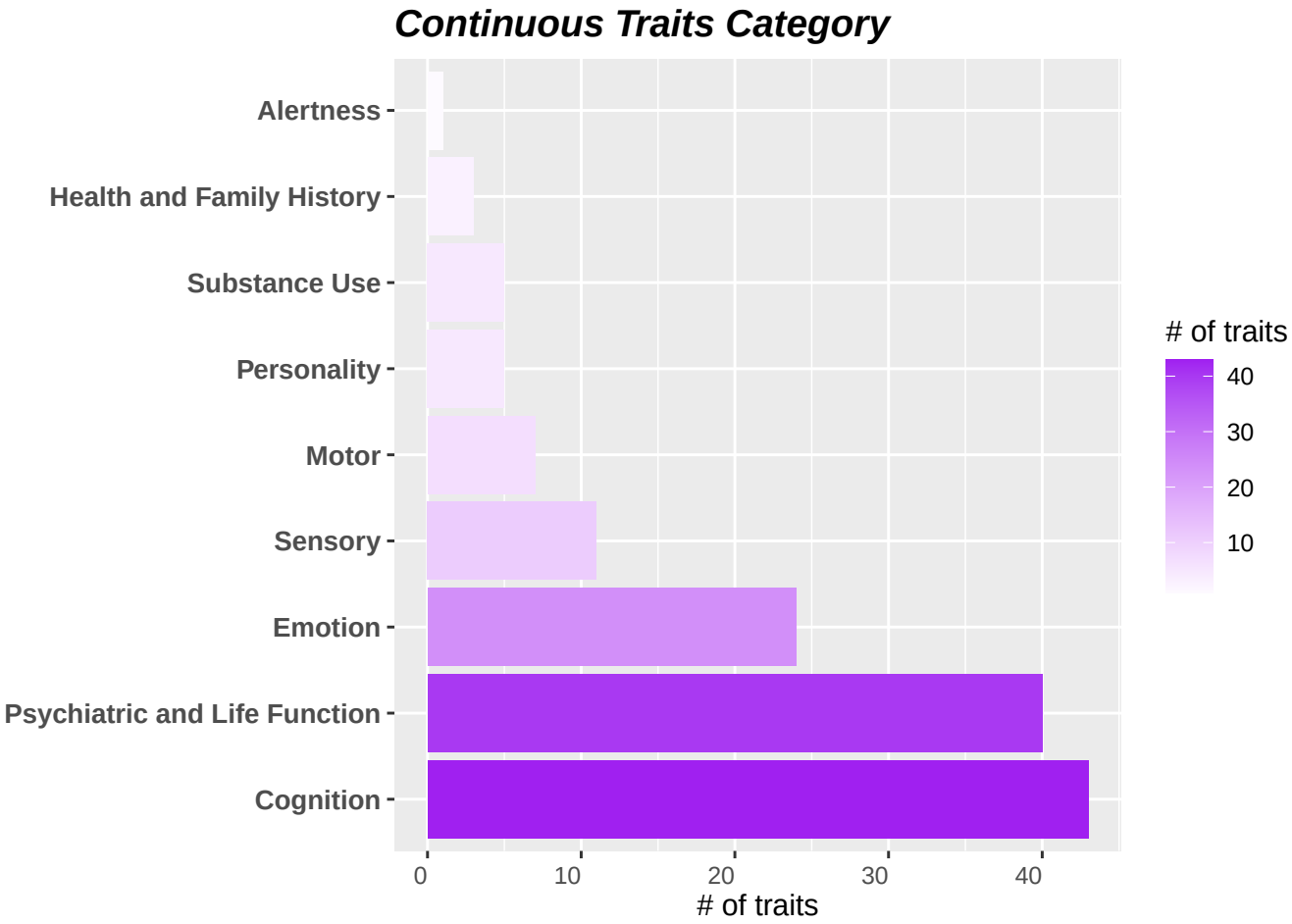
Network Embedding



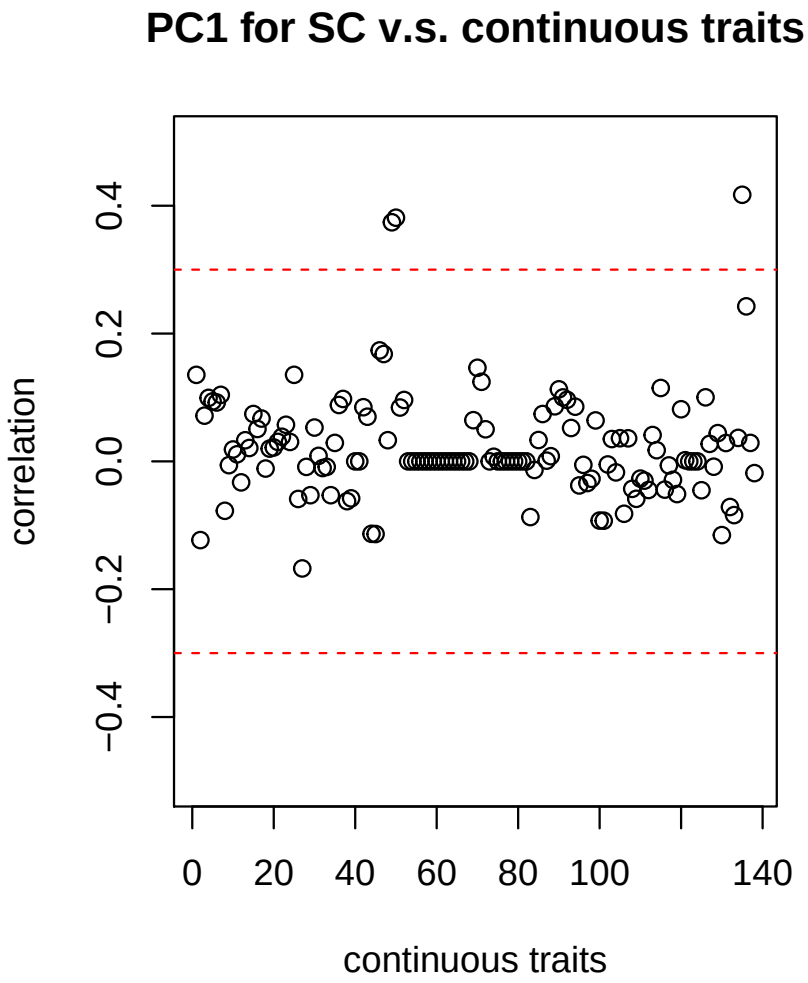


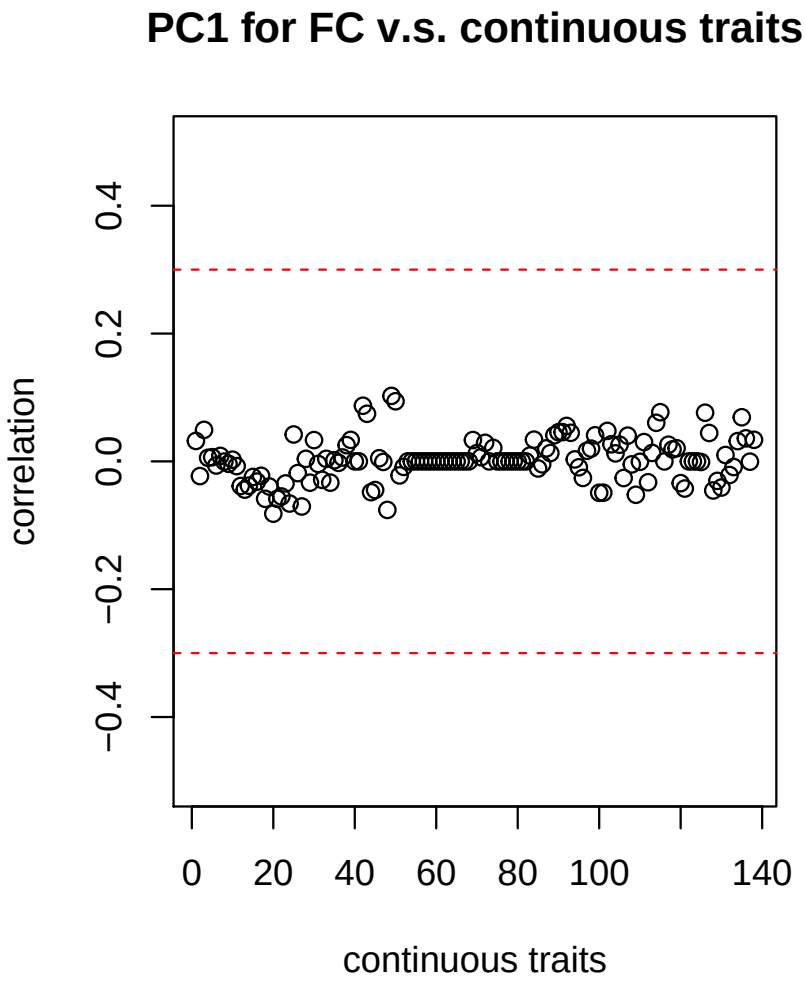
Outline

Continuous Traits

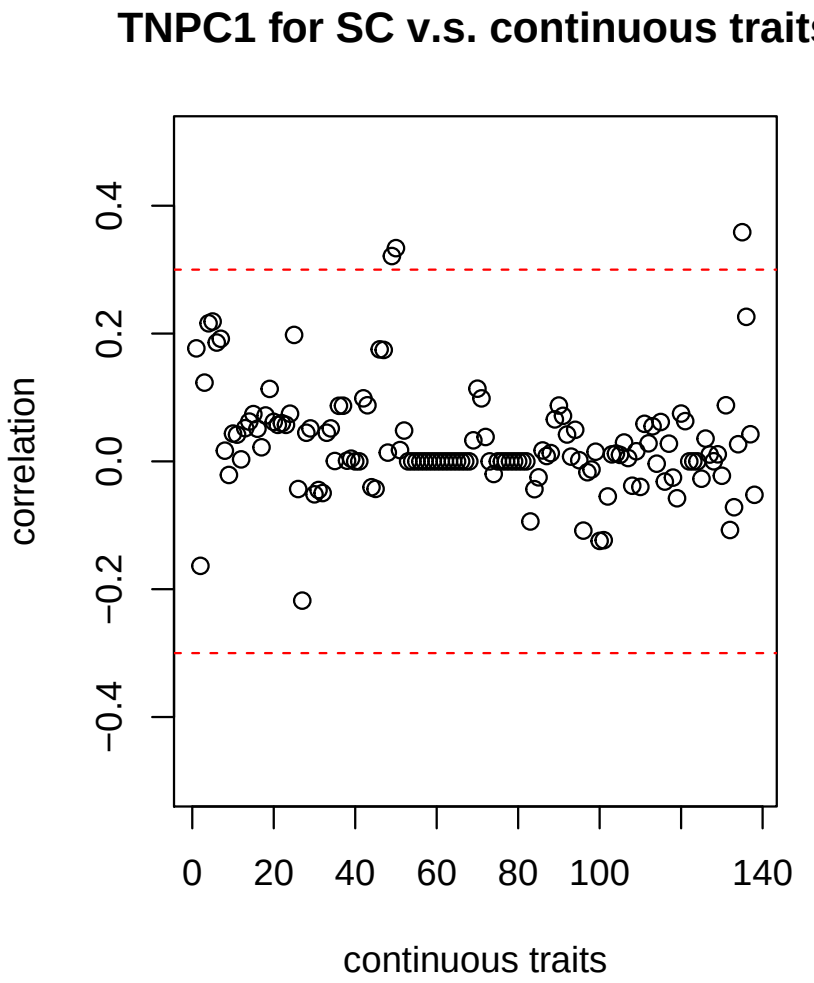


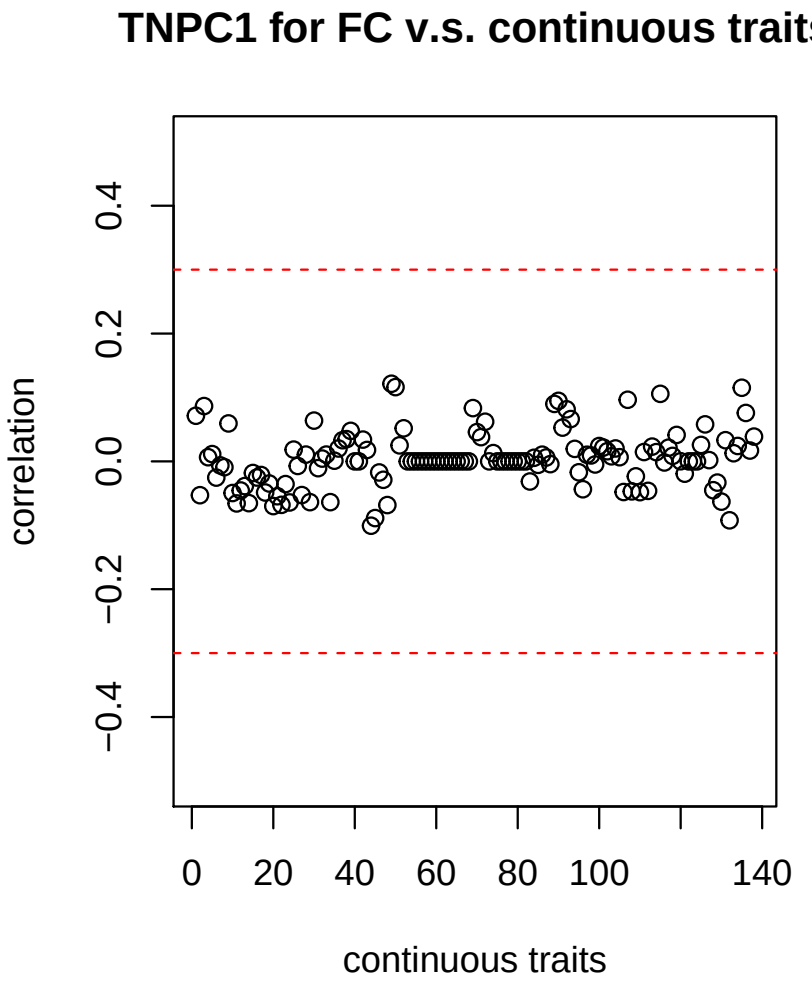
Correlation btw PC1 and continuous traits





Correlation btw TNPC1 and continuous traits

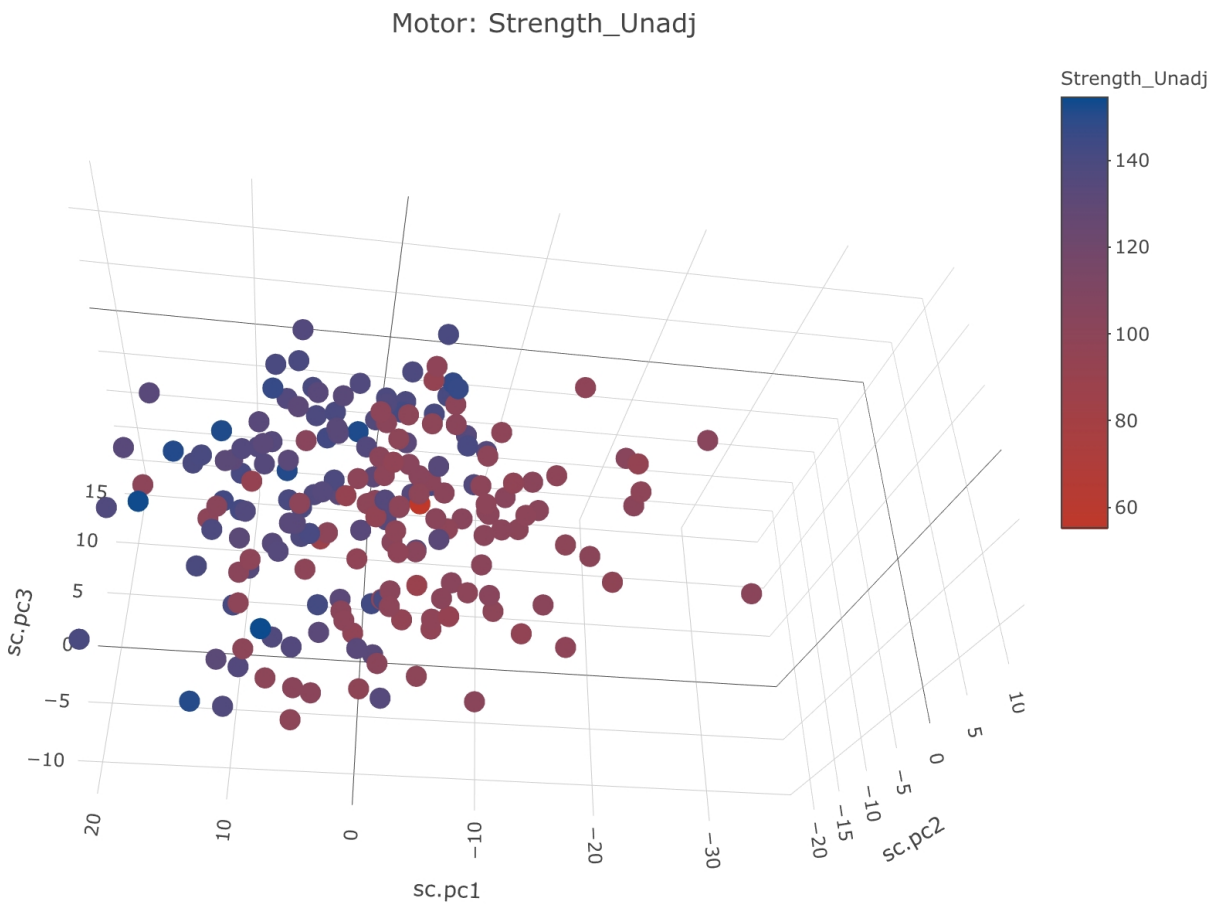




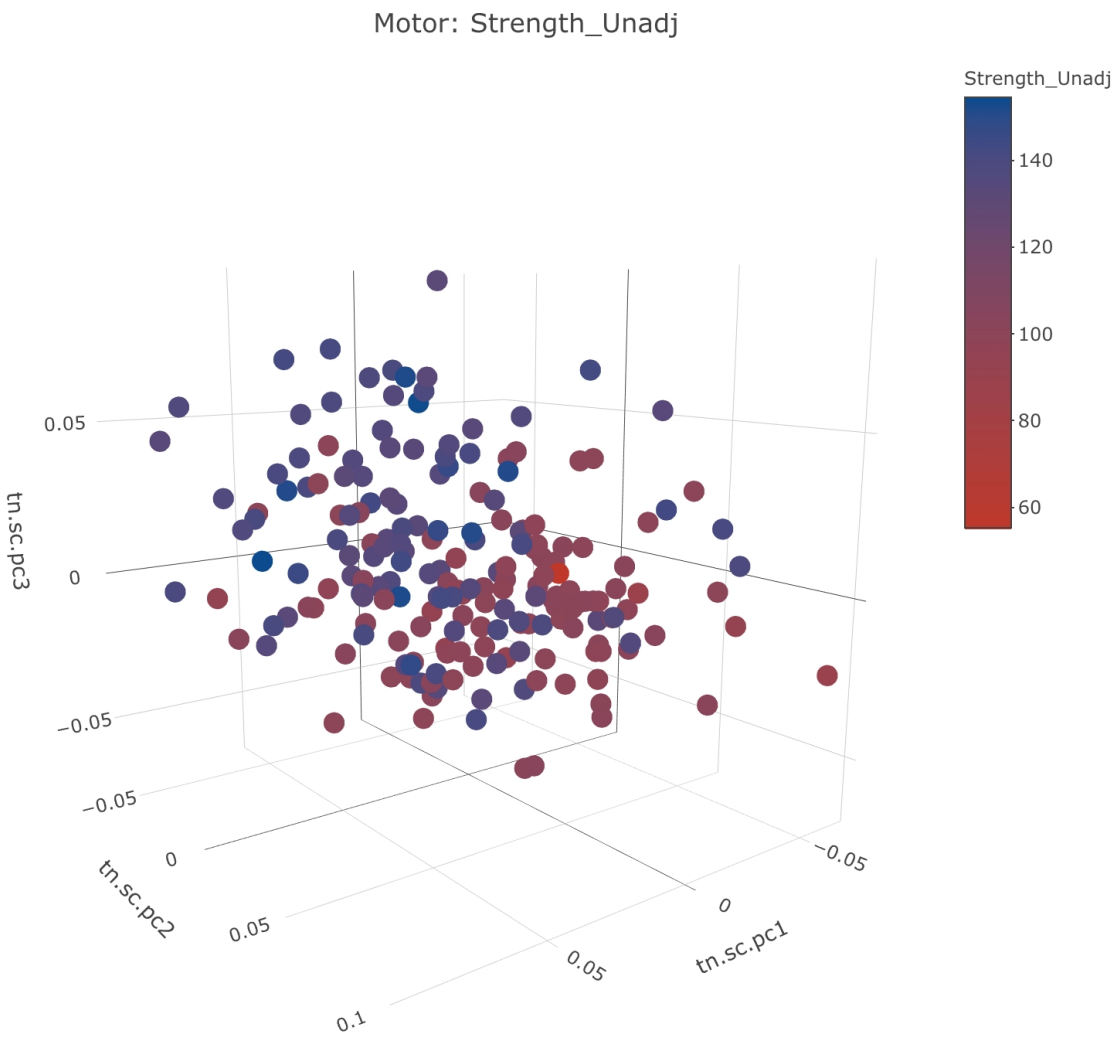
Correlation btw PC1 and continuous traits

- Traits "Strength_Unadj", "Strength_AgeAdj", and "Height" are moderately correlated to the first PC score for structural-related brain network data.
- "Strength_Unadj", "Strength_AgeAdj" are human traits in the motor domain, while "Height" is human trait in the health and family history domain.
- Human traits in the motor domain seem to associate to the structural-related brain network data.

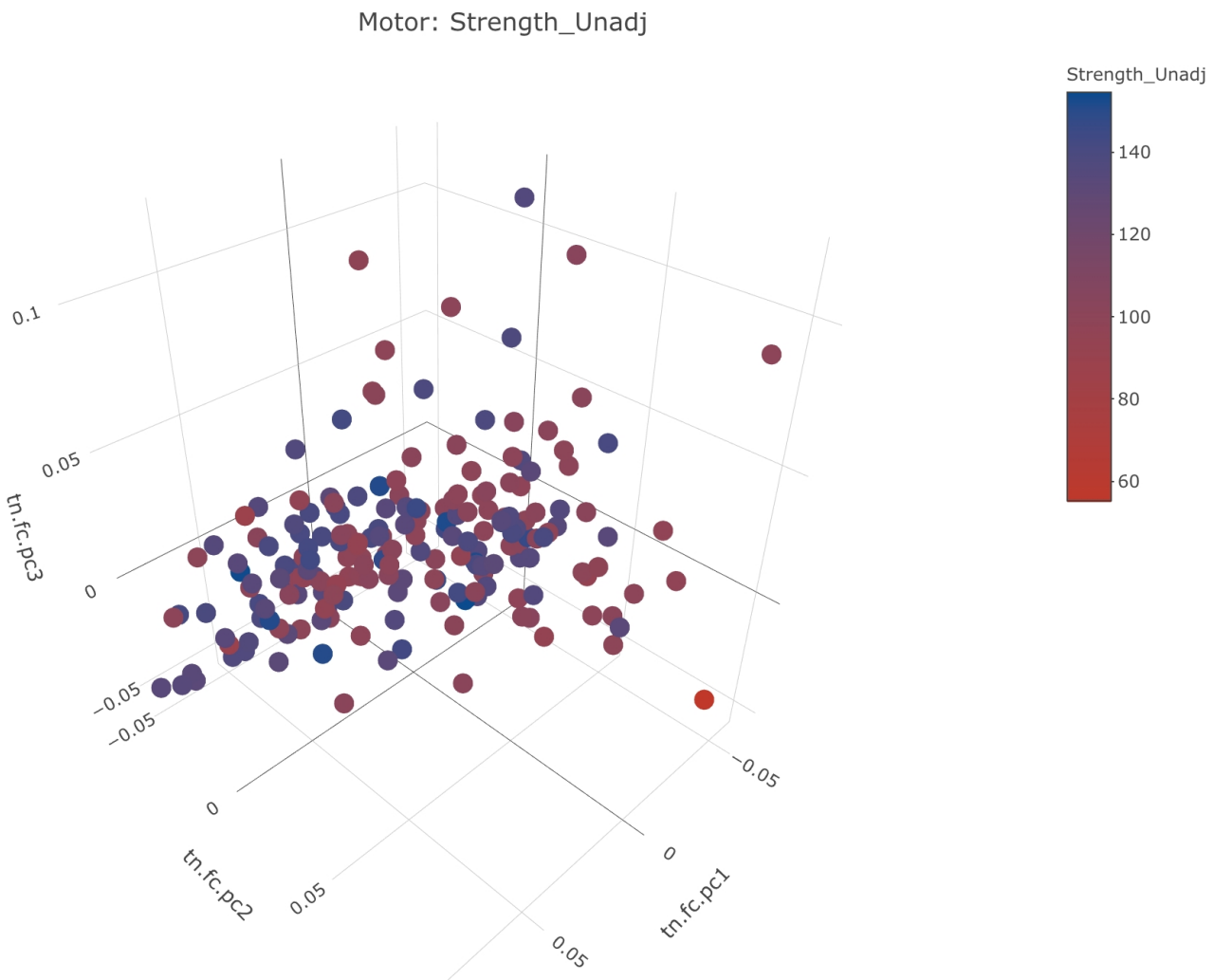
Compare Motor trait with 3 PCA scores on SC



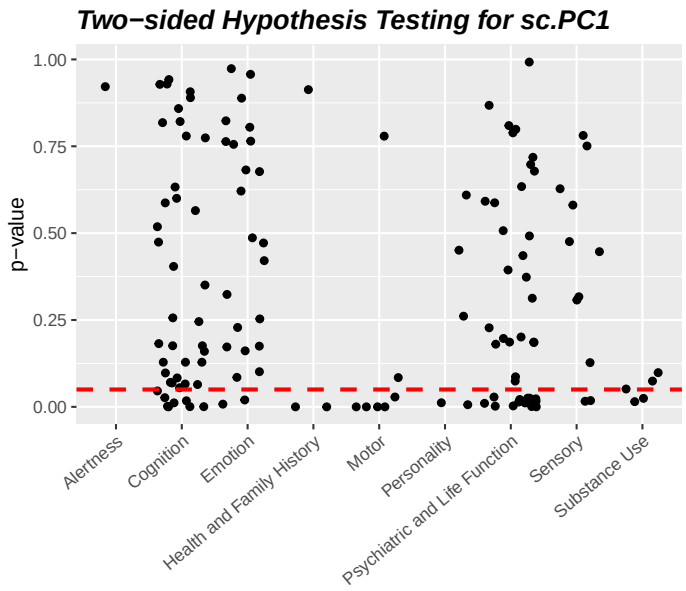
Compare Motor trait with 3 tensor PCA scores on SC

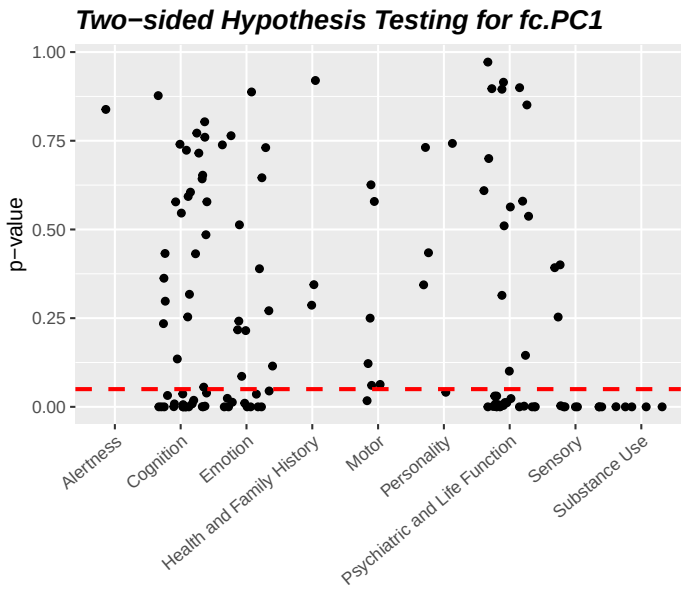


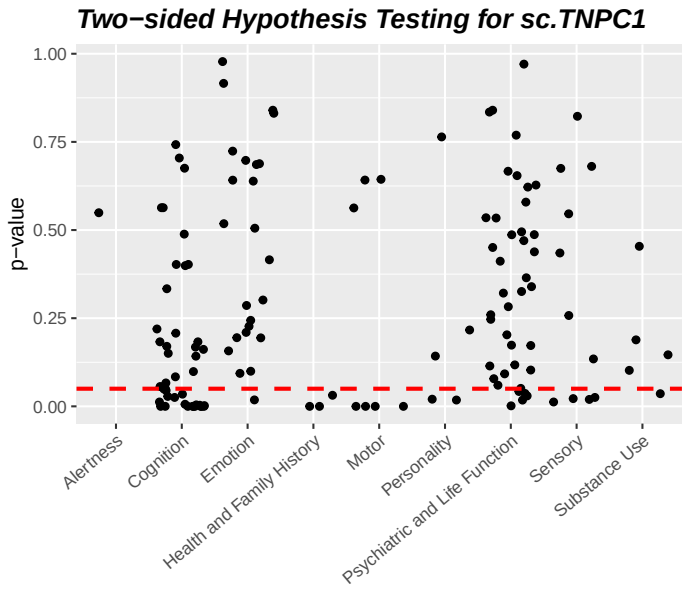
Compare Motor trait with 3 tensor PCA scores on FC

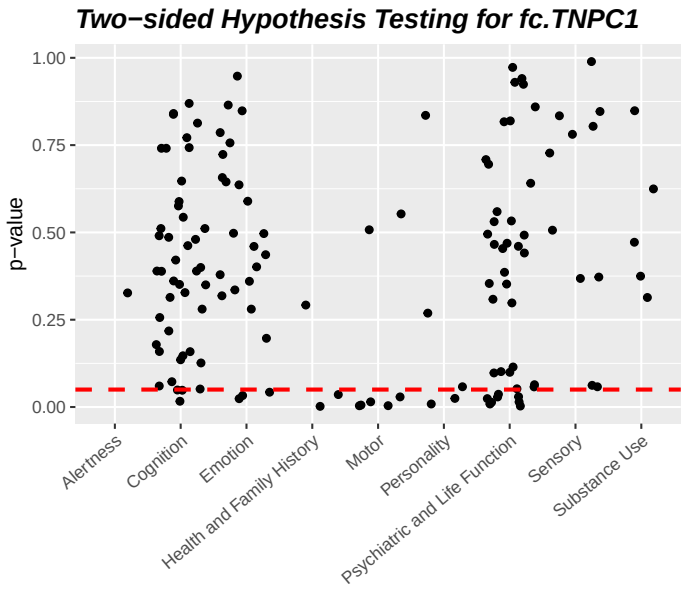


Hypothesis Testing







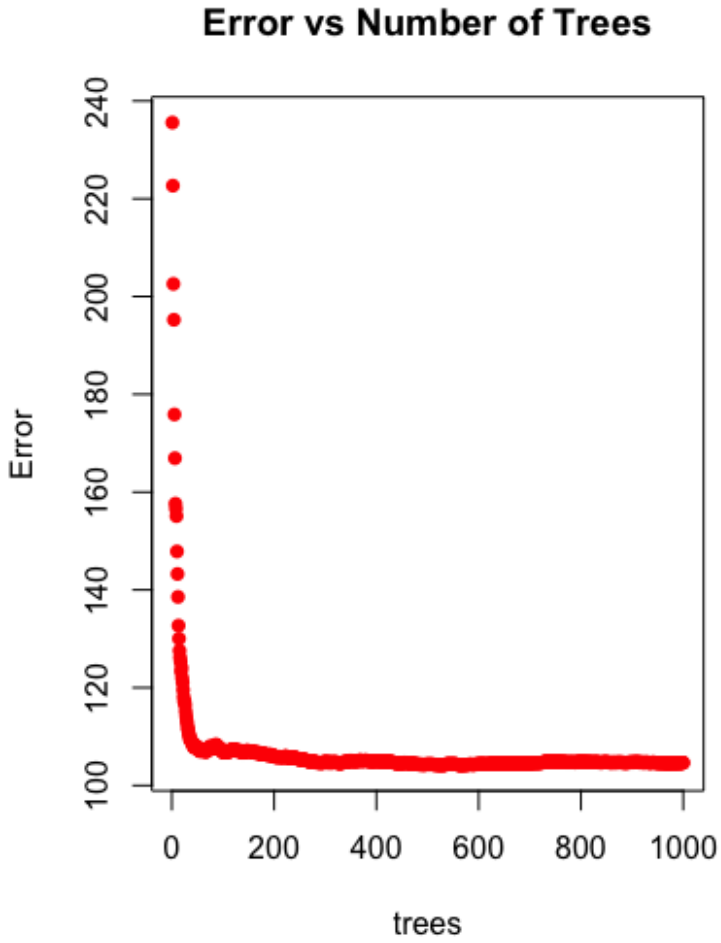


Outline

StrengthUnadj - Regression

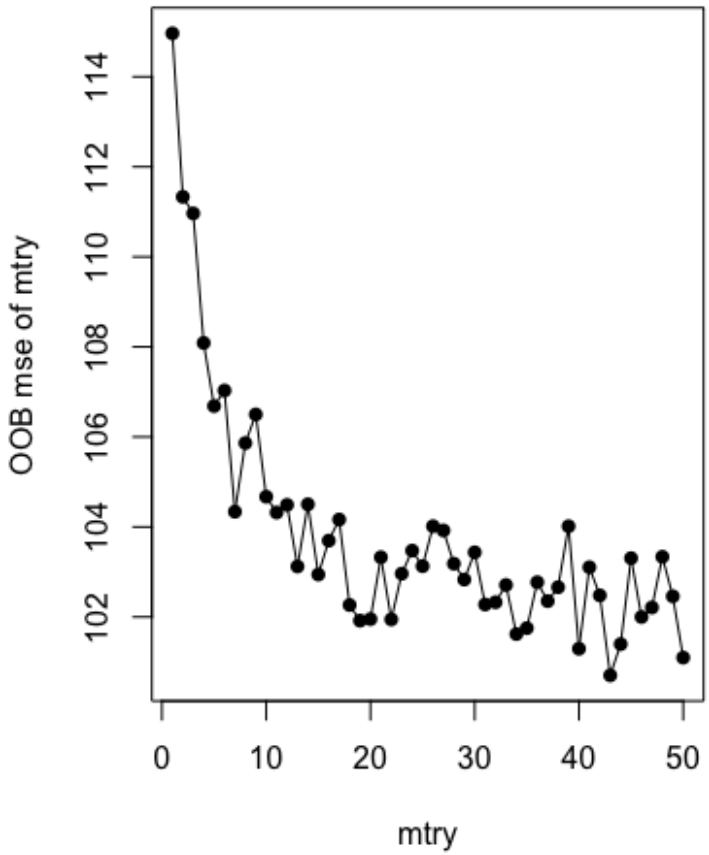
- Linear regression
  - R^2: 0.6391
  - Adjusted R^2: 0.3979
  - Testing error: 11.51053
- Single tree
  - Testing error:11.64323
- Random forest
  - *ntree*
  - *mtry*

Random forest - *ntree*



Set $mtry = 10$ and $ntree = 1000$. Error doesn't decrease when the number of trees is larger than 300 so 300 trees is enough.

Random forest - *mtry*



Let *mtry* be 43 since it minimizes error.

StrengthUnadj - Regression

- Linear regression
  - R^2:0.6391
  - Adjusted R^2:0.3979
  - Testing error:11.51053
- Single tree
  - Testing error:11.64323
- Random forest
  - *ntree*:300
  - *mtry*: 43
  - Testing error:10.34742

StrengthUnadj - Classification

- KNN
  - K value: square root of number of observations in training data = 26
  - Classification accuracy: 60%
- SVM
  - Classification accuracy (linear kernel): 66%
  - Classification accuracy (radial kernel): 73%
  - Relationship between brain connectivity and trait seems to be nonlinear

Continuous Traits - Regression

Trait	Linear regression	Single tree	Randomforest
Edurance_UnAdj	15.5425501	14.0776366	13.1588653
Edurance_AgeAdj	18.2140049	17.4669208	15.3036013
GaitSpeed_Comp	0.2900603	0.2508767	0.2276338
Strength_Unadj	11.5105328	11.6432316	10.3489604
Strength_AgeAdj	20.3156004	21.0215660	18.6210587

Table: Testing errors using SC and FC data

Testing error of single tree and randomforest are almost the same as testing error of linear regression. Single tree and randomforest don't improve predictive power.

Continuous Traits - Classification

Trait	KNN	SVM (linear)	SVM (radial)
Edurance_UnAdj	0.5880682	0.5454545	0.5965909
Edurance_AgeAdj	0.6335227	0.5511364	0.6051136
GaitSpeed_Comp	0.5977337	0.5920680	0.5779037
Strength_Unadj	0.6051136	0.6562500	0.7272727
Strength_AgeAdj	0.6789773	0.6590909	0.7159091

Table: Classification accuracy using SC and FC data

We have 138 continuous traits in total. Only SVM with radial kernel has a classification accuracy higher than 70% for some of them: Strength_Unadj, Strength_AgeAdj and Height.

Comparison

Trait	SVM (SC and FC data)	SVM (SC data)
Strength_Unadj	0.7272727	0.6988636
Strength_AgeAdj	0.7159091	0.7357955
Height	0.7670455	0.7585227

Table: Classification accuracy comparison

"StrengthUnadj", "StrengthAgeAdj", and "Height" are moderately correlated to the first PC score for structural-related brain network data. Classification accuracy doesn't change much when we only use structural related data.

Summary

- Single tree and random forest doesn't have better predictive power than linear regression.
- SVM with radial kernel has a higher classification accuracy than SVM with linear kernel so the relationship between brain network data and traits might be nonlinear.
- SVM(with radial kernel) has a high classification accuracy (higher than 70 %) on traits "StrengthUnadj", "StrengthAgeAdj" and "Height".
- Classification accuracy on above three traits almost doesn't change when we use structural related data instead of using structural related and functional related data. So these three traits seem to associate to the structural-related brain network data.